Journal of Intelligent and Fuzzy Systems 1 (2018) 1–11
IOS Press

A new possibilistic classifier for mixed categorical and numerical data based on a bi-module possibilistic estimation and the generalized minimum-based algorithm

Karim Baati [a,b,*], Tarek M. Hamdani [a,c] Adel M. Alimi [a] and Ajith Abraham [d]
[a] *REGIM-Lab.: REsearch Groups on Intelligent Machines, University of Sfax, National Engineering School of Sfax (ENIS), BP 1173, Sfax, 3038, Tunisia*
E-mail: karim.baati@enis.tn,{tarek.hamdani,adel.alimi}@ieee.org
[b] *HESTIM School of Engineering and Management, Casablanca, Morocco*
[c] *Taibah University, College Of Science And arts at Al-Ula, Al-Madinah a l-Munawwarah, KSA*
[d] *Machines Intelligence Research Labs (MIR Labs), Scientific Network for Innovation and Research Excellence. P.O. Box 2259, Auburn, WA 98071, USA*
E-mail: ajith.abraham@ieee.org

Abstract. In this paper, we suggest NPC_m, a new Naïve Bayesian-like Possibilistic Classifier for mixed categorical and numerical data. The proposed classifier is based on a bi-module belief estimation as well as the Generalized Minimum-based (G-Min) algorithm which has been recently proposed for the classification of categorical data. Distinctively, in the design of both categorical and numerical belief estimation modules, we make use of a probability-to-possibility transform-based possibilistic approach as a strong alternative to the probabilistic one when dealing with decision-making under uncertainty. Thereafter, we use the G-Min algorithm as an improvement of the minimum algorithm to make decision from possibilistic beliefs. Experimental evaluations on 12 datasets taken from University of California Irvine (UCI) and containing all mixed data, confirm the effectiveness of the proposed new G-Min-based NPC_m. Indeed, with the used datasets, the proposed classifier outperforms all the classical Bayesian-like classification methods. Consequently, we prove the efficient use of the bi-module possibilistic estimation approach together with the G-Min algorithm for the classification of mixed categorical and numerical data.

Keywords: Naïve possibilistic classifier, possibility theory, mixed data, Naïve Bayesian classifier, uncertainty

1. Introduction

Selecting the appropriate classification technique for a decision-making problem is a crucial fact that considerably relies on specifications of input data. Among these specifications, we can find the type of input data which may be either categorical, numerical or mixed. Mixed data, as they imply, reflect data which are formed of both categorical and numerical data [1]. These data may be encountered in several real-world applications such as heart disease diagnosis [2], decision on biodegradability of chemicals [3], credit decision making [4], etc.

Bayesian-like classifiers, namely naïve Bayes and naïve possibilistic classifiers stand for straightforward classifiers that assume the independence of input features. Despite their strong assumption, Bayesian-like classifiers can often outperform more sophisticated classification models [5]. Furthermore, two versions may be found for each Bayesian-like classifier:

*Corresponding author. E-mail: karim.baati@enis.tn

one for discrete attributes and another for numerical ones. Therefore, three main approaches may cope with mixed data using Bayesian-like classifiers. In the first approach, an adequate algorithm such as the Entropy-MDLP algorithm (Minimum Description Length Principle) [6] is called in order to discretize the numerical part of data. Later, the Bayesian-like classifier version for categorical data is used in order to estimate beliefs from the obtained outcomes. In the second approach, numerical values are assigned to the categorical part of information (for instance for a given attribute "temperature", we assign the number 1 to the value "high", the number 2 to the value "intermediate" and the number 3 to the value "low"). Afterward, we make use of the appropriate version for numerical data to estimate beliefs. Finally, in the third approach, we maintain the categorical data categorical and the continuous data continuous and we estimate beliefs using the adequate Bayesian-like classifier version for each subset of data. Intuitively, the third approach is likely to be the most efficient since it keeps data in their raw type and hence avoids additional transformation.

In recent studies, naïve possibilistic classifier has attracted more attention among Bayesian-like classification models. In particular, various naïve possibilistic classifiers which are based on the probability-possibility transformation rule of Dubois et al. [7] have been proposed to deal with decision-making from imperfect data. Indeed, a classifier based on this transformation in the discrete case and calling a new decision algorithm called the Generalized minimum-based (G-Min) algorithm has been successfully applied on categorical data [8]. Further, two classifiers based on the same transformation in the continuous case and dealing with numerical features, namely, Naïve Possibilistic Classifier for numerical data and Flexible Naïve possibilistic classifier for numerical data, have been efficiently investigated on numerical information [9].

This study is motivated by the emergence of new probability-to-possibility transform-based possibilistic classifiers which have shown good performance when dealing with uncertainty from either categorical [10][11][8] or numerical [9][12][13] data. In this context, we propose NPC_m, a Naïve Possibilistic Classifier for mixed data as a novel technique that can cope with mixed categorical and numerical information. On the one hand, the proposed classifier maintains the raw type of each data subset (categorical and numerical) by using a bi-module estimation. On the other hand, it makes use of the G-Min algorithm that has proven its effectiveness when classifying categorical data.

The remainder of the paper is structured as follows. Related work is reported in Section 2. Next, in Section 3, we restate the theoretical foundations of the possibilistic classification. Afterward, Section 4 details different technical aspects of the proposed G-Min-based NPC_m. The experimentation results are communicated in Section 5. Lastly, Section 6 concludes the paper and suggests some directions for future research.

2. Related work

A first version of Bayesian-like possibilistic classifier for categorical data was proposed in [15] in order to cope with imprecise training data. However, no efficient method was proposed in this work for the elicitation of possibilistic distributions. Indeed, it was stated in [15] that the procedure proposed to estimate possibility distributions which is based on the computation of the maximum-based projection [14], may construct pathological cases [15].

Later, Haouari et al. [16] have proposed a naïve possibilistic network classifier which models imperfect attribute values when no prior knowledge is available. In this particular context, possibility distributions are built using the partial ignorance which is expressed by an expert.

In [17], authors proposed an efficient algorithm which makes use of the Jeffrey's rule in order to reconsider possibilistic knowledge included by a naïve product-based possibilistic network classifier on the basis of uncertain inputs. The proposed algorithm presents the main advantage of being able to process the classification task in polynomial time with respect to the number of features.

In [18], authors suggested a new Bayesian classifier for uncertain categorical or continuous attributes which integrates uncertainty in the Bayesian theorem and makes use of a new parameter estimation method. Moreover, in [19], authors developed a classification algorithm which can generate rules from uncertain continuous data. For the two works in [18] and [19], authors use intervals in order to model uncertainty over continuous attribute values.

In [9], authors proposed two kinds of possibilistic classifiers for numerical data. In the first one, authors extended the classical and the flexible Bayesian classifiers by applying a probability-possibility transformation to Gaussian distributions. In the second one, authors directly expressed data in possibilistic formats using the idea of proximity between data values. Experiments in [9] have shown the good performance of the flexible possibilistic classifier when dealing with numerical data. Later, authors in [21] extended the probability-to-possibility transform based classifiers which were proposed in [9] in order to deal with uncertainty and imprecision in data modeling. Experimental results reported in [21] have confirmed the robust behavior of the probability-to-possibility transform-based classifiers when they treat imperfect data.

In [8], authors suggested NPC_c, a new naïve possibilistic classifier for categorical data. The proposed classifier relies on the possibilistic approach to estimate beliefs from categorical data and makes use of the G-Min as a novel algorithm to make decision from possibilistic beliefs. Experimental results in [8] have proved the efficiency of the possibilistic approach together with the G-Min algorithm for the classification of categorical data.

In a more recent study, three improved naïve possibilistic classifiers have been proposed in [20]. The new techniques aim to classify imprecise data effectively by relaxing two strong assumptions, namely attributes independence and their equal importance.

Based on previous studies, we can notice that Bayesian-like possibilistic classifiers which have been proposed in the literature are either suitable for categorical or numerical data. By contrast, we report in the current paper a new Bayesian-like possibilistic method which is convenient to classify mixed categorical and numerical data.

3. Possibilistic classification : theoretical background

In order to recall some basics of the possibilistic framework, the following notations are considered in this section as well as in the remainder of this paper:

$C = \{c_1, c_2, ..., c_j, ..., c_C\}$: an exhaustive and exclusive universe of discourse of classes

$A = \{a_1, a_2, ..., a_i, ..., a_M\}$: a set of attributes which may stand for either categorical or numerical values.

v_{it} : the value taken by a given attribute a_i during test.

3.1. Possibility theory

Possibility theory, introduced by Zadeh [22] and then developed by Dubois and Prade [23] is a fusion theory based on fuzzy sets theory and devoted to represent and combine imperfect information in a qualitative or quantitative way. Information imperfections treated by possibility theory may represent the uncertainty due to variability of observations, the uncertainty due to incomplete information, the information ambiguity, the information imprecision, etc [24].

At the semantic level, the basic function in possibility theory is a possibility distribution denoted as π which assigns to each possible class c_j from C a value in [0, 1]. The possibility value assigned to a class c_j stands for plausibility i.e. the belief degree that this class is the right one.

By convention, $\pi(c_j) = 1$ means that c_j is totally possible and if $\pi(c_j) = 0$, c_j is considered as impossible. Intermediary values distinguish values which are more possible than others. Finally, note that in the normalization version of a possibility distribution, we require that at least one class of C is totally possible.

On the other hand, as being based on fuzzy sets theory, possibility theory may recall all the panoply of fuzzy combination rules in order to fuse possibility estimates of a given class [25].

3.2. Conditional Possibility

Possibilistic conditioning corresponds to revising an initial possibility distribution when a new information becomes available. In possibility theory, conditioning may be defined through a counterpart of the Bayes rule. In fact, it can be stated for two subsets E and F of 2^C by:

$$\Pi(E \cap F) = \Pi(E|F) * \Pi(F) \tag{1}$$

where * is a combination operator which is commonly selected as the minimum or the product [26].

3.3. Naïve Bayes Style Possibilistic Classification

Similarly to Bayesian classification (see Appendix A for a reminder), possibilistic classification relies on the aforementioned possibilistic version of the Bayes rule and can be defined by:

$$\pi(c_j|a_1 = v_{1t}, ..., a_M = v_{Mt}) = \frac{\pi(a_1 = v_{1t}, ..., a_M = v_{Mt}|c_j) * \pi(c_j)}{\pi(a_1 = v_{1t}, ..., a_M = v_{Mt})} \tag{2}$$

By assuming that there is no a priori knowledge about classes and the input vector to be classified, we can take $\pi(c_j) = 1$ and $\pi(a_1 = v_{1t}, ..., a_M = v_{Mt}) =$

1. Moreover, as in naïve Bayesian combination, naïve possibilistic classification assumes that attributes $\{a_i\}$ ($\forall\ 1 \leq i \leq M$) are independent. In this case, the conditional joint possibility $\pi(a_1 = v_{1t}, ..., a_M = v_{Mt}|c_j)$ is equal to the fusion of conditional possibility estimations stemming from each single attribute a_i. Therefore, in a context characterized by no a priori knowledge about classes and input vector as well as independent attributes, equation 4 becomes [15] :

$$\pi(c_j|a_1 = v_{1t}, ..., a_M = v_{Mt}) = \pi(a_1 = v_{1t}|c_j) * ... * \pi(a_M = v_{Mt}|c_j) \quad (3)$$

As stated for the conditioning rule, * may be taken as either the product or the minimum [26].

In practice, given a new instantiation $\{a_1 = v_{1t}, ..., a_M = v_{Mt}\}$, we must establish a matrix Π of possibilistic estimates in order to perform the product-based or the minimum-based classification. This matrix is defined as follows :

$$\Pi = \{\pi(i,j)\} = \{\pi(a_i = v_{it}|c_j)\} \quad (4)$$

$\forall\ 1 \leq i \leq M$ and $\forall\ 1 \leq j \leq C$

Lastly, the final decision stands for the class c^* for which equation 3 yields the highest degree of possibility:

$$c^* = arg\ \max_j(\pi(1,j) * ... * \pi(M,j)) \quad (5)$$

4. Proposed Generalized Minimum-based Possibilistic Classifier for mixed data (G-Min-based NPC$_m$)

In order to make decision from mixed categorical and numerical data, we propose a new possibilistic classifier which rests on two main steps as illustrated in Fig. 1.

In the first step, a bi-module belief estimation is applied on mixed data. In order to achieve that, we make use of a specific estimation module that is convenient to each type of data. Indeed, for categorical data, we call the estimation module related to the Naïve Possibilistic classifier for categorical data (NPC_c) which has been proposed in [8]. As for numerical data, we make use of the estimation module related to the Naïve Possibilistic Classifier for numerical data (NPC_n) which has been introduced in [9].

Finally, the G-Min algorithm is employed in order to make decision from the obtained possibilistic estimates. This algorithm stands for an improvement of the classical minimum-based classification algorithm and it aims to improve the classification quality by finding a more reliable final decision.

In the following, technical details of both the bi-module estimation and the G-Min algorithm are given.

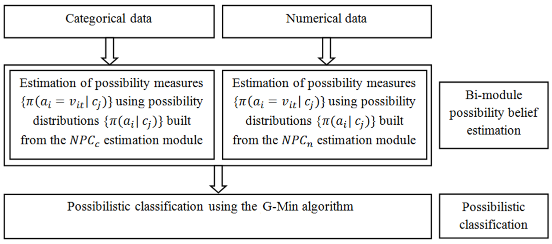

Fig. 1. General structure of the G-Min-based NPC_m

4.1. Bi-module estimation

As mentioned earlier, the bi-module estimation is based on respectively the two estimation modules of NPC_c and NPC_n. Each of the two modules relies on the probabilitiy-to-possibility transformation of Dubois et al. in respectively the discrete and continuous case [7].

4.1.1. Estimation module for categorical data

To illustrate the estimation module for categorical data, let consider $V_i = \{v_{i1}, v_{i2}, ..., v_{is}, ..., v_{iz}\}$ the set of possible categorical values for a given attribute a_i. In order to build possibility distributions from V_i, we start by computing the conditional probability measures $\{p(a_i = v_{is}|c_j)\}$ ($\forall\ 1 \leq s \leq z$) for each attribute a_i with respect to each class c_j using the maximum likelihood estimation (see details in Appendix A). Later, we use the probability-possibility transformation of Dubois et al. in the discrete case [7] in order to express possibilistic estimates. This transformation is defined by :

$$\pi(a_i = v_{is}|c_j) = \sum_{v_{ir}|p(v_{ir}) \leq p(v_{is})} p(a_i = v_{ir}|c_j) \quad (6)$$

with $1 \leq r \leq z$

4.1.2. Estimation module for numerical data

To estimate possibilistic beliefs from numerical data, attribute values are first normalized as follows:

$$a_{in} = \frac{a_i - min(a_i)}{max(a_i) - min(a_i)} \quad (7)$$

where min and max represent respectively the minimum and the maximum value of the attribute a_i over the training set. For simplicity reasons, we consider that all attribute value a_i's used in NPC_n estimation module [9] as well as in all estimation modules for numerical data are normalized and then refer to the corresponding values a_{i_n}.

As in the Naïve Bayes Classifier for numerical data (NBC_n) (see Appendix A), it has been assumed in the NPC_n [9] that each attribute a_i is a random variable which is normally distributed over each class c_j. Thus, a Gaussian distribution $g_{ij} = g(a_i, \mu_{ij}, \sigma_{ij})$ is computed for each feature a_i with regard to each class c_j. For this Gaussian, parameters μ_{ij} and σ_{ij} represent, respectively, the mean and the standard deviation of the variable a_i for the class c_j and they are estimated from training samples. Based on the obtained Gaussian, authors of [9] give the estimation of the probability $P(I_{a_i})$ of the confidence interval centered at μ_{ij} as:

$$P(I_{a_i}|c_j) = 2 * G(a_i, \mu_{ij}, \sigma_{ij}) - 1 \quad (8)$$

where G is a Gaussian cumulative distribution which may be assessed using the table of the standard normal distribution. Later, authors of [21] call the probability-possibility transformation of Dubois et al. in the continuous case [7] in order to estimate the possibility $\pi(a_i = v_{it}|c_j)$ of each test value as follows [9]:

$$\begin{aligned}\pi(a_i = v_{it}|c_j) &= 1 - (2 * G(v_{it}, \mu_{ij}, \sigma_{ij}) - 1) \\ &= 2 * (1 - G(v_{it}, \mu_{ij}, \sigma_{ij}))\end{aligned} \quad (9)$$

4.2. Generalized-minimum-based possibilistic classification algorithm (G-Min algorithm)

As mentioned in Section 3, the minimum-based classification algorithm assigns the final decision to the class that satisfies:

$$c^* = arg \max_j (\min_i \pi(i,j)) \quad (10)$$

where $\Pi = \{\pi(i,j)\}$ ($1 \leq i \leq M$ and $1 \leq j \leq C$) the matrix of possibilistic estimates which has been defined in equation 4.

In many cases, the final class issued from the minimum-based possibilistic classification may have very close possibility estimate to other alternatives. In such situation, the final class may not be the right one and therefore the quality of decision is likely to be seriously altered. In this context, the G-Min has been proposed in order to avoid the ambiguity between the final decision and the rest of hypotheses and hence to find a decision with a possibility estimate widely away from other alternatives (with at least a preset threshold value α). Indeed, this algorithm reflects a "wise" behavior that delays the final decision until a reliable one is got.

Technically speaking, the G-Min algorithm requires the matrix Π of possibilistic estimates and is based on two main steps. The first aims to establish a set of possible decisions whereas the second aims to filter this set in order to find a reliable final class.

More details about the proposed algorithm may be found in Algorithm 1 as well as in [8].

5. Experimental evaluations

This section provides experimental results for the proposed G-Min-based NPC_m and for a set of Bayesian-like classifiers that have been applied on mixed data. This set of classifiers includes three major subsets:

- Naïve Bayesian-like classifiers for categorical data : this subset includes Naïve Bayes Classifier for categorical data (NBC_c) (See Appendix A) and NPC_c. For this subset of classifiers, the numerical part of data is discretized using the Entropy-MDLP (Minimum Description Length Principle) algorithm [6].
- Naïve Bayesian-like classifiers for numerical data : this subset includes Naïve Bayes Classifier for numerical data (NBC_n) (See Appendix A), Flexible Naïve Bayes Classifier for numerical data ($FNBC_n$) (See Appendix A), NPC_n and Flexible Naïve Possibilistic Classifier for numerical data ($FNPC_n$) (See Appendix B). For this subset of classifiers, we have assigned numerical values to the categorical part of inputs.
- Naïve Bayesian-like classifiers for mixed data : this subset involves classifiers with the same pattern of NPC_m but in which the bi-module estimation is differently designed. These classifiers

Algorithm 1 G-Min Algorithm

Require: Matrix of possibilistic estimations Π
Ensure: Final decision c^*

$\Pi \leftarrow sort(\Pi,'rows','increase')$
$\Pi \leftarrow sort(\Pi,'columns','decrease')$
$q \leftarrow 1$
$dset[q] \leftarrow arg(\pi(1,1))$
$c^* \leftarrow arg(\pi(1,1))$
for $j \leftarrow 2\ to\ C$ **do**
  if $\pi(1,1) - \pi(1,j) < \alpha$ **then**
    $q \leftarrow q+1$
    $dset[q] \leftarrow arg(\pi(1,j))$
  end if
end for
$card \leftarrow cardinality(dset)$
$i \leftarrow 2$
while $card \neq 1\ and\ i \neq A$ **do**
  for $q \leftarrow 1\ to\ dset.size$ **do**
    $poss[q] \leftarrow arg^{-1}(dset[q], i)$
  end for
  $max \leftarrow maximum(poss)$
  for $q \leftarrow 1\ to\ poss.size$ **do**
    if $max - poss[q] \geq \alpha$ **then**
      $dset \leftarrow dset.remove(arg(poss[q])$
      $c^* \leftarrow max$
    end if
  end for
  $card \leftarrow cardinality(dset)$
  $i \leftarrow i+1$
end while

are namely, Naïve Bayes Classifier for mixed data (NBC_m), Flexible Naïve Bayes Classifier for mixed data ($FNBC_m$) and Flexible Naïve Possibilistic Classifier for mixed data ($FNPC_m$). Description of the bi-module estimation in each of these classifiers is given in Table 1.

Table 1

Description of the bi-module estimation in Bayesian-like classifiers for mixed data

Classifier	Categorical data	Numerical data
NBC_m	NBC_c estim. module	NBC_n estim. module
$FNBC_m$	NBC_c estim. module	$FNBC_n$ estim. module
$FNPC_m$	NPC_c estim. module	$FNPC_n$ estim. module

During the experimental part of our work, we have used the Bayes rule to make decision from probabilistic beliefs whereas the product, the minimum as well as the G-Min rules have been investigated in order to carry out the possibilitic classification. Moreover, we have taken $\alpha = 0.1$ in order to perform the G-Min-based classification.

Experiments are conducted on 12 datasets from UCI machine learning repository [27] containing all mixed data. Characteristics of these datasets are illustrated in Table 2. As shown in Table 2, the number of cases ranges from 57 to 3136, the number of attributes from 9 to 41 and the number of classes from 2 to 7. Moreover, some datasets contain missing values and others no. Therefore, a wide variety of problems is represented.

Table 2

Details of datasets used in the experiments

Dataset	Instances	Missing values?	Att.	Cat.	Num.	Classes
Sick	3163	Yes	25	18	7	2
Contra.	1473	No	9	7	2	3
QSAR	1055	No	41	4	37	2
German	1000	No	20	13	7	2
Credit	690	Yes	15	9	6	2
Cylinder	512	Yes	39	19	20	2
Horse	368	Yes	22	15	7	2
Heart	270	No	13	7	6	2
Auto	205	Yes	25	10	15	7
Flags	194	No	28	18	10	6
Hepat.	155	Yes	19	13	6	2
Labor	57	No	16	8	8	2

During experiments, we have carried a 10-cross validation and as in [9], we have used the standard Percent of Correct Classification (PCC) to assess performances of different Bayesian-like classifiers. PCC is defined by:

$$PCC = \frac{number\ of\ well\ classified\ instances}{total\ number\ of\ instances} * 100 \quad (11)$$

Experimental results obtained with various Bayesian-like classifiers are shown in Table 3.

Table 3

Experimental results obtained with various Bayesian-like classifiers given as the mean and the standard deviation

	Sick	Contra.	QSAR	German	Credit	Cylinder	Horse	Heart	Auto	Flags	Hepat.	Labor	Average rank
NBC$_c$	95.89 (3)	53.78 (3)	78.58 (8)	**75.90 (1)**	85.91 (4)	71.29 (14)	80.29 (6)	82.59 (10)	71.92 (17)	71.00 (7)	84.90 (11)	93.02 (7)	6.74
NPC$_c$ (Prod.)	95.77 (4)	50.92 (9)	78.77 (7)	68.90 (9)	84.46 (7)	73.33 (13)	79.48 (8)	82.80 (9)	72.66 (16)	71.66 (4)	85.66 (7)	92.78 (8)	8.41
NPC$_c$ (Min.)	95.66 (5)	47.87 (15)	77.54 (11)	62.40 (19)	81.44 (10)	70.74 (15)	79.80 (7)	78.59 (15)	72.91 (14)	71.66 (4)	85.88 (6)	92.22 (9)	10.83
NPC$_c$ (G-Min.)	96.19 (2)	49.57 (13)	83.50 (5)	66.80 (13)	86.07 (2)	74.81 (9)	80.63 (4)	82.48 (11)	86.33 (2)	71.97 (3)	88.29 (2)	95.66 (2)	5.82
NBC$_n$	84.22 (20)	50.18 (11)	75.23 (13)	68.50 (10)	77.97 (16)	75.92 (8)	78.74 (9)	83.70 (6)	57.50 (20)	57.63 (14)	83.82 (12)	93.63 (5)	11.99
FNBC$_n$	88.35 (19)	54.36 (2)	64.93 (20)	70.00 (8)	80.40 (14)	81.66 (2)	80.39 (5)	82.96 (8)	61.32 (19)	58.36 (13)	85.24 (8)	93.45 (6)	10.33
NPC$_n$ (Prod.)	89.88 (17)	53.00 (4)	66.34 (17)	66.40 (14)	69.24 (18)	78.33 (3)	64.62 (19)	84.04 (4)	73.33 (12)	69.07 (10)	85.00 (10)	83.33 (19)	12.24
NPC$_n$ (Min.)	92.07 (13)	46.03 (17)	82.66 (6)	65.10 (17)	80.84 (11)	70.18 (17)	65.18 (16)	77.03 (16)	75.66 (8)	49.18 (18)	81.92 (16)	88.00 (13)	14.00
NPC$_n$ (G-Min.)	93.88 (8)	51.83 (7)	74.60 (14)	70.30 (7)	85.04 (6)	76.85 (7)	80.72 (3)	81.85 (12)	**87.73 (1)**	55.97 (16)	86.24 (4)	**96.66 (1)**	7.16
FNPC$_n$ (Prod.)	94.11 (7)	**55.04 (1)**	83.98 (2)	71.80 (4)	85.97 (3)	**82.48 (1)**	81.04 (2)	83.33 (7)	84.66 (4)	56.31 (15)	86.15 (5)	93.66 (4)	4.58
FNPC$_n$ (Min.)	92.47 (9)	49.70 (12)	73.74 (15)	65.80 (15)	67.21 (19)	74.44 (10)	72.00 (14)	76.92 (17)	86.03 (3)	47.52 (20)	77.20 (20)	90.00 (11)	13.74
FNPC$_n$ (G-Min.)	92.35 (10)	50.88 (10)	77.82 (10)	70.90 (5)	80.70 (12)	73.70 (12)	72.22 (13)	84.22 (3)	75.00 (10)	73.78 (2)	83.13 (14)	85.00 (17)	9.83
NBC$_m$	90.35 (16)	51.42 (8)	78.09 (9)	75.60 (2)	80.55 (13)	74.22 (11)	65.72 (15)	84.44 (2)	75.24 (9)	71.12 (6)	85.15 (9)	84.92 (18)	9.83
FNBC$_m$	88.40 (18)	45.56 (19)	83.70 (4)	65.60 (16)	85.18 (5)	70.55 (16)	78.74 (9)	80.24 (13)	76.00 (7)	47.86 (19)	83.27 (13)	88.33 (12)	12.58
NPC$_m$ (Prod.)	92.17 (12)	48.68 (14)	75.64 (12)	68.40 (11)	82.15 (8)	77.59 (5)	77.40 (11)	83.98 (5)	68.00 (18)	51.55 (17)	83.02 (15)	91.22 (10)	11.50
NPC$_m$ (Min.)	90.42 (14)	51.98 (6)	65.77 (19)	70.70 (6)	69.39 (17)	78.14 (4)	65.09 (17)	80.00 (13)	73.66 (11)	69.60 (9)	86.40 (3)	83.33 (20)	11.58
NPC$_m$ (G-Min.)	**97.69 (1)**	52.01 (5)	**84.24 (1)**	73.60 (3)	**88.02 (1)**	77.02 (6)	**82.84 (1)**	**87.92 (1)**	82.33 (5)	**74.81 (1)**	**88.72 (1)**	93.95 (3)	**2.41**
FNPC$_m$ (Prod.)	94.74 (6)	47.00 (16)	66.20 (18)	66.90 (12)	78.08 (15)	55.74 (20)	75.75 (12)	73.59 (19)	80.00 (6)	64.52 (11)	79.46 (18)	96.98 (15)	14.00
FNPC$_m$ (Min.)	90.42 (14)	45.85 (18)	72.80 (16)	62.65 (18)	64.75 (20)	58.03 (19)	63.82 (20)	69.85 (20)	73.33 (12)	62.78 (12)	77.50 (19)	85.27 (14)	16.83
FNPC$_m$ (G-Min.)	92.22 (11)	45.24 (20)	83.89 (3)	62.00 (20)	81.65 (9)	70.00 (18)	64.88 (18)	75.81 (18)	72.88 (15)	70.15 (8)	81.81 (17)	85.12 (16)	14.41

From Table 3, we can see that:

- The proposed classifier is the best in terms of average rank among all the 20 compared Bayesian-like classifiers.
- The proposed classifier is the best-ranked on 7 among the 20 benchmarks (namely, in datasets Sick, QSAR, Credit, Horse, Heart, Flags and Hepatitis).
- The second most efficient classifier among the set of classifiers is the product-based $FNCP_n$. That confirms the good performance of this classifier which has been already reported in [9] and justifies the legitimacy of the use of a probability-to-possibility transform-based method for possibility estimation within the proposed classifier.
- The proposed G-Min-based NPC_m outperforms the minimum-based NPC_m and the product-based NPC_m as well. These results reveal that the efficiency of the proposed G-Min-based NPC_m is not due to the only use of the possibilitic estimation approach but rather to the simultaneous use of this approach with the G-Min algorithm.

On the other side, in order to compare the proposed classifier with the rest of Bayesian-like classifiers in terms of PCC, we have used the Wilcoxon Matched-Pairs Signed-Ranks Test as detailed in [28]. It is a non-parametric test which is devoted to compare two classifiers over multiple data sets.

Comparison results using the Wilcoxon Matched-Pairs Signed-Ranks Test are given in Table 4 and they show that the proposed G-Min-based NPC_m outperforms all the compared Bayesian-like classifiers ($p - value < 0.05$).

6. Conclusion

The appeal of the proposed work is related to the need to involve mixed data in real-world applications and hence the necessity to find techniques which perform well when they handle this type of data.

In this context, we have proposed the G-Min-based NPC_m as a new technique which makes decision from mixed data pervaded with uncertainty. Based on experimentation and comparison results with a panoply of Bayesian-like classifiers, we have proven the efficiency of the suggested classifier. That can be explained by the fact that the possibilistic approach in both estimation (bi-module possibilistic method) and classification (G-Min algorithm) rests on fuzzy logic which is a strong framework to deal with decision making under uncertainty in real world settings [29].

To conclude, the current work is likely to be considered, from a deeper point of view, as a study that aimed to select the best classifier that deals with mixed data among a particular family of classifiers (Bayesian-like). Thus, many interesting studies in the future may be conducted with the same objective to find the most accurate classifier for mixed data among other classification families such as logic-based classification methods, perceptron-based techniques and Support Vector Machines (SVM) [30]. Indeed, only based on such studies (in addition to the current one) that we can find the classification technique which is absolutely the best when handling mixed categorical and numerical information.

Acknowledgments

The authors would like to acknowledge the financial support of this work by grants from General Direction of Scientific Research (DGRST), Tunisia, under the ARUB program.

References

[1] A. R. De Leon and K. C. Chough (Eds.), *Analysis of Mixed Data: Methods & Applications*, CRC Press, 2013.

[2] J. Nahar, T. Imam, K. S. Tickle and Y. P. P. Chen, Computational intelligence for heart disease diagnosis: A medical knowledge driven approach, *Expert Systems with Applications* **40(1)** (2013), 96–104.

[3] K. Mansouri, T. Ringsted, D. Ballabio, R. Todeschini and V. Consonni, Quantitative structure-activity relationship models for ready biodegradability of chemicals, *Journal of chemical information and modeling* **53(4)** (2013), 867–878.

[4] J. Li, L. Wei, G. Li and W. Xu, An evolution strategy-based multiple kernels multi-criteria programming approach: The case of credit decision making, *Decision Support Systems* **51(2)** (2011), 292–298.

[5] P. Langley, W. Iba and K. Thompson, *An analysis of Bayesian classifiers*, in: Proceedings of AAAI, 1992, vol. 90, pp. 223–228.

[6] U. M. Fayyad and K. B. Irani, On the handling of continuous valued attributes in decision tree generation, *Machine learning* **8(1)** (1992), 87–102.

[7] D. Dubois, L. Foulloy, G. Mauris and H. Prade, Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities, *Reliable computing* **10(4)** (2004), 273–297

[8] Baati, K., Hamdani, T.M., Alimi, A.M., Abraham, A.: A New Classifier for Categorical Data Based on a Possibilistic Estimation and a Novel Generalized Minimum-based Algorithm. Journal of Intelligent and Fuzzy Systems 33(3), 1723–1731 (2017)

Table 4

Results for the Wilcoxon Matched-Pairs Signed-Ranks Test

NPC_m(G-Min.) Vs NBC_c	**NPC_m(G-Min.) Vs NPC_c(Prod.)**	**NPC_m(G-Min.) Vs NPC_c(Min.)**
$p<0.0386$	$p<0.0004$	$p<0.0004$
NPC_m(G-Min.) Vs NPC_c(G-Min.)	**NPC_m(G-Min.) Vs NBC_n**	**NPC_m(G-Min.) Vs $FNBC_n$**
$p<0.0386$	$p<0.0004$	$p<0.0386$
NPC_m(G-Min.) Vs NPC_n(Prod.)	**NPC_m(G-Min.) Vs NPC_n(Min.)**	**NPC_m(G-Min.) Vs NPC_n(G-Min.)**
$p<0.0063$	$p<0.0004$	$p<0.0004$
NPC_m(G-Min.) Vs $FNPC_n$(Prod.)	**NPC_m(G-Min.) Vs $FNPC_n$(Min.)**	**NPC_m(G-Min.) Vs $FNPC_n$(G-Min.)**
$p<0.0386$	$p<0.0004$	$p<0.0386$
NPC_m(G-Min.) Vs NBC_m	**NPC_m(G-Min.) Vs $FNBC_m$**	**NPC_m(G-Min.) Vs NPC_m(Prod.)**
$p<0.0386$	$p<0.0386$	$p<0.0063$
NPC_m(G-Min.) Vs NPC_m (Min.)	**NPC_m(G-Min.) Vs $FNPC_m$(Prod.)**	**NPC_m(G-Min.) Vs $FNPC_m$(Min.)**
$p<0.0004$	$p<0.0063$	$p<0.0004$
NPC_m(G-Min.) Vs $FNPC_m$ (G-Min.)		
$p<0.0063$		

[9] M. Bounhas, K. Mellouli, H. Prade and M. Serrurier, Possibilistic classifiers for numerical data, *Soft Computing* **17** (2013), 733–751.

[10] K. Baati, T. M. Hamdani and A. M. Alimi, *Diagnosis of lymphatic diseases using a naïve bayes style possibilistic classifier*, in: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC), IEEE, 2013, pp. 4539–4542.

[11] Baati, K., Hamdani, T.M., Alimi, A.M., Abraham, A.: A Modified Naïve Bayes Style Possibilistic Classifier for the Diagnosis of Lymphatic Diseases. In: Proceedings of the 16th International Conference on Hybrid Intelligent Systems, Springer, 2016, pp. 479–488.

[12] Baati, K., Hamdani, T.M., Alimi, A.M., Abraham, A.: A Modified Naïve Possibilistic Classifier for Numerical Data. In: Proceedings of the 16th International Conference on Intelligent Systems Design and Applications, Springer, 2016, pp. 417–426.

[13] Baati, K., Hamdani, T.M., Alimi, A.M., Abraham, A.: Decision Quality Enhancement in Minimum-based Possibilistic Classification for Numerical Data. In: Proceedings of the 8th International Conference on Soft Computing and Pattern Recognition, Springer, 2016, pp. 353–358.

[14] C. Borgelt and J. Gebhardt, *A naïve bayes style possibilistic classifier*, in: Proceedings of the 7th European Congress on Intelligent Techniques and Soft Computing, 1999.

[15] C. Borgelt and R. Kruse, *Efficient maximum projection of database induced multivariate possibility distributions*, in: Proceedings of the 7th IEEE international conference on fuzzy systems, 1988, pp. 663–668.

[16] B. Haouari, N. Ben Amor, Z. Elouedi and K. Mellouli, Naïve possibilistic network classifiers, *Fuzzy Sets and Systems* **160(22)** (2009), 3224–3238

[17] S. Benferhat and K. Tabia, *An efficient algorithm for naïve possibilistic classifiers with uncertain inputs*, in Proceedings of the 2nd International Conference on Scalable Uncertainty Management (SUM), Springer LNAI, 2008, vol. 5291, pp. 63–77.

[18] B. Qin, Y. Xia, and F. Li, *A Bayesian classifier for uncertain data*, in the 25th ACM Symposium on Applied Computing (SAC), 2010, pp. 1010–1014.

[19] B. Qin, Y. Xia, S. Prabhakar, and Y. Tu, *A rule-based classification algorithm for uncertain data*, in IEEE International Conference on Data Engineering, 2009, pp. 1633–1640.

[20] J. Bai, Y. Yang, and J. Xie, Improved Naïve Possibilistic Classifiers for Imprecise Data. *IAENG International Journal of Computer Science* **45(1)** (2018), 153–163.

[21] M. Bounhas, M. G. Hamed, H. Prade, M. Serrurier and K.Mellouli, Naïve possibilistic classifiers for imprecise or uncertain numerical data, *Fuzzy Sets and Systems* **239** (2014), 137–156.

[22] L. A. Zadeh, Fuzzy sets as a basis for a theory of possibility, *Fuzzy sets and systems* **1(1)** (1978), 3–28.

[23] D. Dubois and H. M. Prade, *Possibility theory: an approach to computerized processing of uncertainty*, New York: Plenum press, 1988.

[24] B. Khaleghi, A. Khamis, F. O. Karray and S. N. Razavi, Multisensor data fusion: A review of the state-of-the-art, *Information Fusion* **14(1)** (2013), 28–44.

[25] K. Baati, T.M. Hamdani and A.M. Alimi, *A modified hybrid naïve possibilistic classifier for heart disease detection from heterogeneous medical data*, in: Proceedings of the 6th International Conference on Soft Computing and Pattern Recognition, IEEE, 2014, pp. 353–35.

[26] D. Dubois and H. Prade, The logical view of conditioning and its application to possibility and evidence theories, *International Journal of Approximate Reasoning* **4(1)**(1990), 23–46.

[27] J. Mertz and P. M. Murphy, *UCI repository of machine learning databases*, Available at: ftp://ftp.ics.uci.edu/pub/machine-learning-databases.**160** (2007).

[28] J. Demsar, Statistical comparisons of classifiers over multiple data sets, *The Journal of Machine Learning Research* **7** (2006), 1–30.

[29] L.A. Zadeh, A note on similarity-based definitions of possibility and probability, *Information Sciences* **267** (2014), 334–336.

[30] S. B. Kotsiantis, I. D. Zaharakis and P. E. Pintelas, Supervised machine learning: A review of classification techniques, *Frontiers in Artificial Intelligence and Applications*

[31] D. Geiger and D. Heckerman, *Learning gaussian networks*, in: Proceedings of the Tenth International Conference on Uncertainty in Artificial Intelligence, 1994, pp. 235–243.

[32] G.H. John, P. Langley, *Estimating continuous distributions in Bayesian classifiers*, in: Proceedings of the 11th conference on uncertainty in artificial intelligence, 1995.

Appendix A: Naïve Bayesian Classifier

Naïve Bayes classifier rests on the Bayes rule in order to compute the posterior probability for each class c_j in the presence of a new instantiation $\{a_1 = v_{1t}, ..., a_M = v_{Mt}\}$. Bayes rule is defined by:

$$p(c_j|a_1 = v_{1t}, ..., a_M = v_{Mt}) = \frac{p(a_1 = v_{1t}, ..., a_M = v_{Mt}|c_j).p(c_j)}{p(a_1 = v_{1t}, ..., a_M = v_{Mt})} \quad (12)$$

The term $P(a_1 = v_{1t}, ..., a_M = v_{Mt})$ is a normalization factor which can be ignored. Moreover, in naïve Bayesian classification settings, attributes are assumed to be independent so that the Bayes rule becomes:

$$p(c_j|a_1 = v_{1t}, ..., a_M = v_{Mt}) = \prod_{i=1}^{M} p(a_i = v_{it}).p(c_j) \quad (13)$$

When dealing with categorical data, naïve Bayes classifier computes probability distributions from training data using the maximum likelihood estimation. Indeed, in the Naïve Bayesian Classifier for categorical data (NBC_c), the probability of each discrete value v_{is} ($\forall\ 1 \leq s \leq z$) of a given attribute a_i is defined by:

$$p(a_i = v_{is}|c_j) = \frac{\#(a_i = v_{is}, c_j)}{\#c_j} \quad (14)$$

where $\#(a_i = v_{is}, c_j)$ is the number of training samples belonging to the class c_j and having the value v_{is} for the attribute a_i and $\#c_j$ is the number of training samples that belong to the class c_j.

In naïve Bayes classifiers for numerical data, two main approaches may be found in order to build probability distributions. The first (which is called the normal approach) is used within the Naïve Bayes Classifier for numerical data (NBC_n) whereas the second (which is called the kernel approach) is used within the so-called Flexible Naïve Bayes Classifier for numerical data ($FNBC_n$) [32].

In the normal method, each attribute a_i is assumed to be normally distributed over each hypothesis c_j and hence it is supposed to follow a Gaussian distribution as stated in the following :

$$p(a_i|c_j) = g(a_i, \mu_j, \sigma_j) \quad (15)$$

For this Gaussian, parameters μ_j and σ_j represent, respectively, the mean and the standard deviation of training instances belonging to the class c_j. Therefore, in order to build the NBC_n estimation module, μ_j and σ_j are first found from training samples and the probability measure assigned to each test value v_{it} of a given attribute a_i is given as follows [31]:

$$p(a_i = v_{it}|c_j) = \frac{1}{\sqrt{2\pi}.\sigma_j} \exp(-\frac{(v_{it} - \mu_j)^2}{2.\sigma_j^2}) \quad (16)$$

In the kernel method, the normality assumption is abandoned and a nonparametric kernel density formed of a set of Gaussians is used instead. To illustrate this method, let consider N_{ij} the number of values $\{v_{il}\}$ $(1 \leq l \leq N_{ij})$ of a given a_i which are encountered during training with respect to a given class c_j and let σ_j the standard deviation defined by:

$$\sigma_j = \frac{1}{\sqrt{N_j}} \quad (17)$$

where N_j represents the total number of training instances belonging to the class c_j.

In order to build the kernel density, Gaussians $\{g(a_i, v_{il}, \sigma_j)\}$ $(1 \leq l \leq N_{ij})$ are stored during training. Then, the probability measure assigned to each test value v_{it} of a given attribute a_i is given as an average of these Gaussians [32]:

$$p(a_i = v_{it}|c_j) = \frac{1}{N_j} \sum_{l=1}^{N_j} g(v_{it}, v_{il}, \sigma_j) \quad (18)$$

Lastly, the final decision in Bayesian classification stands for the class c^* with the highest posterior probability:

$$c^* = arg\ \max_j (p(c_j|a_1 = v_{1t}, ..., a_M = v_{Mt})) \quad (19)$$

Appendix B: Flexible Naïve Possibilistic Classifier for numerical data (FNPC$_n$)

As in $FNBC_n$, the estimation module of the Flexible Naïve Possibilistic Classifier for numerical data ($FNPC_n$) [16] is based on a set of possibility distributions which are estimated using training samples. In order to illustrate the method used for estimation, we must first consider the standard deviation σ defined in [9] by:

$$\sigma = \frac{1}{\sqrt{N}} \tag{20}$$

where N represents the total number of training instances.

Later, the probability measure assigned to each new value v_{it} of a given attribute a_i is computed as an average of possibility distributions [9]:

$$p(a_i = v_{it}|c_j) = \frac{1}{N_j} \sum_{l=1}^{N_j} \pi(v_{it}, c_{jl}) \tag{21}$$

with:

$$\pi(v_{it}|c_{jl}) = 2 * (1 - G(v_{it}, v_{il}, \sigma)) \tag{22}$$

where N_j represents the total number of training instances belonging to the class c_j, G a Gaussian cumulative distribution which may be assessed using the table of the standard normal distribution and $\{v_{il}\}$ $(1 \leq l \leq N_{ij})$ the training instances of a given a_i with regard to a given class c_j.